

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра социальной информатики

**БАЗОВЫЕ ВОЗМОЖНОСТИ SPSS (НА ПРИМЕРЕ МАССИВА  
ДАНЫХ ПО РЕЗУЛЬТАТАМ АНКЕТИРОВАНИЯ)**

Автореферат бакалаврской работы

Студента 4 курса 451 группы  
направления 09.03.03 – «Прикладная информатика»  
профиль Прикладная информатика в социологии  
Социологического факультета СГУ  
Лесина Ильи Вячеславовича

Научный руководитель  
доктор социологических наук,  
профессор

\_\_\_\_\_ В.Л. Шабанов  
подпись, дата

Зав. кафедрой  
кандидат социологических наук, доцент

\_\_\_\_\_ И.Г. Малинский  
подпись, дата

Саратов 2021

## ВВЕДЕНИЕ

**Актуальность.** Точная и своевременная информация о том, что может произойти в экономике и обществе в будущем, всегда имела значение для тех, кто принимает решения. Прогнозирование стало важной частью процесса планирования стратегии любой компании и политики любого государства. Развитие современных экономических, социологических теорий, а также сложных компьютерных программ повлияло на подъем новых методов прогнозирования и анализа.

В современных условиях, когда информационные потоки стали особенно массивными, появляется колоссальное количество данных. Также, ввиду того, что скорость всех процессов в обществе возрастает, увеличивается и потребность в выявлении быстрых ответов на возникающие вопросы, с целью поиска которых и проводится огромное количество исследований и сбор данных во всех сферах жизни.

Сегодня рынок статистического программного обеспечения отличается многообразием и, одновременно с этим, высокой специализацией. Существует более тысячи разнообразных программ, решающих задачи статистического анализа социологических данных. Однако даже на рынке такого специфического программного обеспечения существует конкуренция.

Актуальность выбора соответствующего инструментария для обработки социологических данных обусловлена тем, что при проведении любых исследований, включающих в себя анализ массивов данных, будь то психологические, медицинские исследования, научные изыскания, а также социологические опросы и исследования, выбор инструментария для их обработки, диагностики и прогнозирования является очень важным для принятия управленческих решений, а на первый план выходят совокупность быстроты, точности и удобства в процессе анализа полученных результатов. В рамках данной дипломной работы будет проведен анализ программного пакета для обработки статистических данных IBM SPSS Statistics.

SPSS является модульной программой для обработки статистических

данных. Ее основу составляет базовый модуль – SPSS Base – позволяющий осуществлять управление данными и содержащий наиболее распространенные методы статистического анализа данных, такие как: проведение описательной статистики, преобразование данных, обработка категориальных (номинальных) переменных, построение графиков, подготовка отчетов, а также вычисление корреляций и проведение многомерного статистического анализа и прочие функции.

**Объектом исследования** является пакет программного обеспечения SPSS.

**Предметом** выступают актуализированные в рабочем режиме функционально-аналитические характеристики SPSS.

**Эмпирическая база исследования.** В качестве массива социологических данных для демонстрации эвристического потенциала SPSS были использованы данные, собранные в ходе проведенного исследования влияния городских парков и зеленых насаждений на физическое и эмоциональное состояние студентов, проведенное мною в 2019 году среди студентов СГУ (объем выборки 134 респондента).

**Целью** бакалаврской работы является описание и анализ базового эвристического потенциала пакета программного обеспечения для обработки статистических данных SPSS.

Сформулированная цель достигается решением следующим **задач**:

- 1) Выявить особенности программных пакетов обработки статистических данных.
- 2) Проанализировать интерфейс SPSS.
- 3) Ввести понятие массива социологических данных применительно к материалам анкетного опроса.
- 4) Провести базовый статистический анализ на базе данных авторского социологического исследования «Влияние городских парков и зеленых насаждений на физическое и эмоциональное состояние студентов-социологов».
- 5) Провести статистический анализ таблиц зависимости на базе данных

авторского социологического исследования.

- б) Провести визуализацию данных – построить графики основных типов, иллюстрирующие некоторые результаты авторского социологического исследования.

**Структура бакалаврской работы.** Работа состоит из введения, двух глав, заключения, списка использованных источников и приложения с текстом анкеты.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**В первой главе** «Программные пакеты для обработки статистических данных» решаются первые две задачи.

**В параграфе 1.1** «Разновидности программных пакетов» анализируются различные программные пакеты, реализующие обработку статистических данных.

С появлением статистических пакетов исследователь может ограничиться лишь базовыми познаниями в математической статистике: достаточно знать и оперировать статистическими понятиями и уметь правильно выбрать метод анализа.

Большинство представленных на рынке статистических пакетов обладают гибкой модульной структурой, которая может пополняться и расширяться за счет пользовательских модулей, дополнительно закупаемых или находящихся в свободном доступе в Интернете. Подобная гибкость позволяет адаптировать большинство пакетов к потребностям конкретного пользователя.

Универсальные статические пакеты должны удовлетворять следующим требованиям: содержать достаточно широкий набор стандартных статистических методов, быть достаточно простым для быстрого освоения и использования непрофессиональным пользователем, работать с достаточно большими базами данных и отвечать высоким требованиям к вводу, преобразованию и организации хранения данных, осуществлять обмен данными с широко распространенными пакетами и базами данных, иметь обширный набор средств графического представления данных и результатов их анализа,

иметь подробное документационное сопровождение и адаптированную под конечного пользователя справочную систему.

Перечисленным требованиям удовлетворяет статистический пакет для социальных наук SPSS Statistics. Среди других известных и популярных статистических пакетов отметим S, R, STATISTICA, STATGRAPHICS, Minitab.

Наряду с универсальными, существуют профессиональные пакеты (SAS, BMDP), которые отличаются от универсальных тем, что позволяют работать со сверхбольшими объемами данных, применять узкоспециализированные методы анализа, создавать собственную систему обработки данных. Они также отличаются гораздо более высокой стоимостью, чем универсальные пакеты.

Для обработки специфических данных, используемых в отдельных областях деятельности, разработано специализированное программное обеспечение, такое как, например, BioStat (для анализа данных в области биологии и медицины, фармакологии и биохимии), MESOSAUR (анализ временных рядов и построение регрессионных моделей), DATASCOPE (анализ многомерных данных).

**Параграф 1.2** «Интерфейс пакета SPSS» посвящен анализу интерфейса SPSS. Отмечаются действия пользователя при установке пакета, требуемые для этого компьютерные ресурсы.

В программном пакете SPSS есть возможность импортировать данные из текстового файла в форматах csv, txt, dat, tab, в которых данные разделены запятой, или из файлов ряда других форматов, в том числе форматов EXCEL xls и xlsx. Дается описание команд, реализующих импорт данных из файлов различного типа.

SPSS использует «классический» оконный интерфейс: при открытии файла с данными пользователь получает возможность работать в окне данных и окне переменных; в последнем пользователь описывает вводимые переменные – такие их характеристики, как тип, размер и др. Результаты статистического анализа выводятся в диалоговом окне «Вывод».

**Во второй главе** «Пример обработки массива социологических данных средствами пакета SPSS» решаются задачи 3-6.

**В параграфе 2.1** «Понятие массива данных и общая характеристика массива, сформированного по итогам авторского социологического исследования» вводится понятие массива данных и представляется конкретный массив данных.

Массив данных – это совокупность однородных записей, характеризующих какой-либо объект управления или процесс, рассматриваемый как одно целое и упорядоченный таким образом, что его описание с помощью индексов однозначно определяет положение каждого элемента или путь доступа к нему. Размерность массива - количество элементов этого массива.

В SPSS используется двумерный массив данных, представляющий собой таблицу (матрицу), обозначаемую  $A(I,j)$ , где  $i$  указывает номер строки, а  $J$  – номер столбца. Если число строк равно  $n$ , а число столбцов –  $m$ , то размерность массива равна  $m*n$ . Если речь идет о массиве анкетных данных, то они кодируются специальным и достаточно очевидным образом в зависимости от типов вопросов (номинальные простые и альтернативные признаки кодируются по-разному). SPSS имеет удобные средства ввода таких данных. Однако, несмотря на наличие таких средств, массив данных по результатам социологического исследования часто вводится в формате EXCEL (xls,xlsx).

В качестве массива социологических данных для проведения статистических расчетов в SPSS были использованы данные, собранные автором в 2019 году в ходе самостоятельно проведенного исследования влияния городских парков и зеленых насаждений на студентов-социологов. Методом сбора первичной эмпирической информации был избран опрос по простой случайной выборке в форме анкетирования, место проведения исследования – социологический факультет СГУ. Объем выборки – 134 респондента. Генеральная совокупность – студенты социологического факультета СГУ общей численностью 210 человек.

Собранные данные представляют собой заполненные анкеты с ответами на вопросы. После оцифровки был получен двумерный массив данных размерностью (134, 32), где 134 – количество опрошенных респондентов, 32 – количество переменных (вопросов анкеты, среди которых два вопроса относились к номинальной шкале с совместимыми альтернативами, и оцифровывались в виде нескольких дихотомических столбцов; их число совпадало с числом альтернатив-ответов).

**В параграфе 2.2** «Выполнение базового анализа данных в программном пакете SPSS» представлен анализ базовых возможностей SPSS по обработке социологических данных – построение одномерных и двумерных (совместных) распределений с проверкой гипотезы о связи двух признаков, составляющих совместное распределение – таблицу сопряженности.

Описаны встроенные возможности SPSS, используемые для вывода таблиц сопряженности с инструментом частотного анализа, и команды, реализующие данные функции. В параграфе представлены 4 одномерные распределения в форме частотных таблиц, реализующие социально-демографические характеристики респондентов. Дается их описание и на его основе – усредненный социальный портрет участника социологического исследования. Это студент-социолог 1-4 курса, обучающийся по одному из направлений – организация работы с молодежью, прикладная информатика в социологии, социология, социальная работа – со средним возрастом 20,2 года.

**В параграфе 2.3** «Составление таблиц сопряженности социологических данных с использованием программного пакета SPSS» представлены совместные распределения ряда признаков анкетного опроса вместе со статистикой Хи-квадрат, проверяющей статистическую гипотезу о наличии связи.

Показана команда и доступ к ней в главном меню SPSS, которая реализует совместные распределения. Представлены окна вывода с двумя таблицами сопряженности, построенными на основе двух пар признаков – «Номер курса»,

«Отношение к экологической ситуации в парках» и «Номер курса», «Отношение к парковым аттракционам».

Прослеживается четкая обратная зависимость между уровнем образования и восприятием экологической ситуации: чем взрослее студент, тем он критичнее. Для формально-статистической проверки гипотезы о наличии зависимости был использован критерий Хи-квадрат. По его результатам было установлено превышение Хи-квадрат расчетного над соответствующим табличным, что позволило отвергнуть нулевую статистическую гипотезу об отсутствии связи между признаками. Итак, связь между курсом обучения и отношением к экологической ситуации имеется.

По аналогии была проанализирована вторая таблица сопряженности. Было установлено, что взросление и повышение уровня образования характеризуется повышением положительного восприятия аттракционов в городском парке (или, возможно, равнодушным их восприятием). Проверка с помощью критерия Хи-квадрат показала превышение Хи-квадрат расчетного над соответствующим табличным, что позволило отвергнуть нулевую статистическую гипотезу об отсутствии связи между признаками. Итак, связь между курсом обучения и восприятием парковых аттракционов имеется.

**В параграфе 2.4 «Визуализация данных в графическом виде в SPSS»** продемонстрированы средства SPSS для построения диаграмм разных типов.

Диаграмма – это графическое представление данных линейными отрезками или геометрическими фигурами, позволяющее быстро оценить соотношение нескольких величин.

Иногда для оформления диаграмм используется трёхмерная визуализация, представленная на двумерной плоскости, что придаёт диаграмме «объемность».

Программный пакет SPSS поддерживает работу с тремя видами диаграмм – столбчатыми, гистограммами и круговыми.

Столбчатые диаграммы в основном используются для наглядного сравнения полученных статистических данных или для анализа их изменения за определённый промежуток времени. Построение столбчатой диаграммы

заключается в изображении статистических данных в виде вертикальных прямоугольников или трёхмерных прямоугольных столбиков. Каждый столбик изображает величину уровня данного статистического ряда. Все сравниваемые показатели выражены одной единицей измерения, поэтому удаётся сравнить статистические показатели данного процесса. Разновидностями столбчатых диаграмм являются линейные (полосовые) диаграммы. Они отличаются горизонтальным расположением столбиков. Столбчатые и линейные диаграммы взаимозаменяемы, рассматриваемые в них статистические показатели могут быть представлены как вертикальными, так и горизонтальными столбиками. В обоих случаях для изображения величины явления используется одно измерение каждого прямоугольника — высота или длина столбика. Поэтому и сфера применения этих двух диаграмм в основном одинакова.

Достаточно распространённым способом графического изображения структуры статистических совокупностей является секторная, или круговая, диаграмма, так как идея целого очень наглядно выражается кругом, который представляет всю совокупность. Относительная величина каждого значения изображается в виде сектора круга, площадь которого соответствует вкладу этого значения в сумму значений. Этот вид графиков удобно использовать, когда нужно показать долю каждой величины в общем объёме. Сектора могут изображаться как в общем круге, так и отдельно, расположенными на небольшом удалении друг от друга.

Круговая диаграмма сохраняет наглядность только в том случае, если количество частей совокупности диаграммы небольшое. Если частей диаграммы слишком много, её применение неэффективно по причине несущественного различия сравниваемых структур. Недосток круговых диаграмм — малая ёмкость, невозможность отразить более широкий объём полезной информации.

В качестве примера для оценки удобства построения диаграмм в SPSS были составлены круговые диаграммы для показателей пола, возраста, курса и направления обучения опрошенных студентов, и столбиковые диаграммы для показателей отношения к экологической ситуации и отношения к аттракционам

из нашего массива данных. Вывод диаграмм осуществляется непосредственно в окне вывода, с подписанными процентными величинами каждого из секторов, различными цветами для секторов и подписанной легендой.

## **ЗАКЛЮЧЕНИЕ**

Любой программный продукт, реализующий анализ статистической информации и другие прикладные аспекты, связанные с этим, является лишь инструментом в руках статистика. Для социолога использование SPSS – обширного программного продукта с благоприятным интерфейсом, рассчитанным на конечного пользователя, не являющегося программистом, – стало составной частью профессионального стандарта.

Основными особенностями SPSS и других программных пакетов обработки статистических данных являются: модульность, дающая возможность расширения пакета за счет добавления новых программ, реализующих дополнительные статистические процедуры; диалоговый режим работы пользователя с пакетом, при котором пользователь выбирает нужную статистическую процедуру из системы многоуровневого меню, при этом пакетом обеспечивается автоматическая генерация команд на достаточно простом языке; поддержка разных форматов, особенно универсальных текстовых и связанных с электронными таблицами.

Благоприятный пользовательский интерфейс обеспечивает как большую совокупность статистических процедур обработки данных, так и удобные средства ввода данных и средства графического и табличного вывода в нескольких форматах.

Основной формой организации данных в SPSS является массив, представляющий собой таблицу вида «объект»-«признак»; при обработке социологических данных в роли объекта часто выступают анкеты, заполненные респондентами, в роли признака – формализованные вопросы анкеты. Каждый признак имеет определенную размерность в зависимости от использованной шкалы. В частности, для номинальной шкалы это набор целых чисел, сопоставленных с номером ответа.

В качестве массива социологических данных для проведения статистических расчетов в SPSS были использованы данные авторского исследования на тему «Влияние городских парков и зеленых насаждений на физическое и эмоциональное состояние студентов-социологов», проведенного в 2019 г. Методом сбора первичной эмпирической информации был анкетный опрос по простой случайной выборке. Объем выборки – 134 респондента. Генеральная совокупность – студенты социологического факультета СГУ. В процессе обработки данных средствами SPSS были установлены возможности пакета по базовой обработке социологических данных, включающей построение одномерных и совместных распределений, а также средства статистической проверки гипотезы о связи признаков. В частности, на примере двух совместных распределений, реализованных средствами SPSS, с помощью критерия Хи-квадрат было показано наличие прямой зависимости между признаками «Курс обучения» и «Отношением к парковым аттракционам» и обратной зависимости между признаками «Курс обучения» и «Отношением к экологической ситуации в парках». По итогам анализа средствами SPSS была проведена визуализация данных – построены графики двух типов, демонстрирующие построенные распределения.

